图书情报工作
第66卷 第6期 2022年3月

面向出版物内容增强的医学文献信息资源服务模式创新*

翟姗姗 刘星月 陈欢
华中师范大学信息管理学院 武汉 430079

摘 要:[目的/意义]医学文献信息资源是用户学习医学知识、展开医学研究的必要前提,然而现有的医学文献信息资源存在非结构化、呈现形式单一等问题,由此提出医学出版物内容增强方案,以期为用户提供更好的医学资源组织与利用服务。[方法/过程]以语义出版为导向,将 CMeSH 医学主题词融入元素识别、语义映射、语义描述、多维关联流程之中,实现医学传统出版物到医学增强出版物的转化,同时以儿科为应用背景构建医学出版物内容增强实例,验证数字出版内容增强在医学文献资源服务中的可行性。[结果/结论]研究发现,通过内容增强一系列流程,医学传统出版物可转化为医学增强出版物,从而优化用户医学知识学习过程,使用户的高效学习成为可能。

关键词:内容增强 数字出版 医学文献信息资源 CMeSH 语义出版
分类号:G250
DOI:10.13266/j.issn.0252-3116.2022.06.011

1 引言

随着大数据时代的到来,医学信息资源呈现指数化增长,海量的医学信息资源成为用户获取医学知识的负担。事实上,随着医学研究进程的不断加快,数字出版成为医学学术期刊的主要出版方式。但现有的数字出版存在模式单一、文献交互性差、缺乏医学专业内容解读等问题,内容增强为医学文献信息资源现存问题的解决提供了思路。内容增强是数字出版过程中的一种增强行为,语义出版是实现数字出版内容增强的有效途径,内容增强以语义出版为导向,具有语义出版的特征,是丰富出版物知识内容与表现方式的重要方式。本文运用语义出版相关技术,对面向内容增强的医学文献信息资源服务提出新的思路和模型,在保证医学出版物规范性的前提下增强医学出版物的可读性,希望借此提升医学文献的学术出版功能,增强医学文献信息资源的利用程度,进而为用户提供更加系统全面的医学文献信息资源服务。

2 相关研究

为了更加科学有效地开展研究,需要通过文献研究法对文中涉及到的研究对象、研究方法进行总结与分析,数字出版内容增强、医学文献信息资源建设与服务则是本文的研究基础。

2.1 数字出版内容增强相关研究

数字出版是利用数字技术对内容进行编辑加工,并借助计算机网络进行传播的一种出版方式[1]。数字出版内容增强以语义出版为导向,具有语义出版的特征,是语义增强的出版形式[2]。在国外,英国皇家化学学会(RSC)最早对数字出版内容增强展开探索,从单篇论文到通过富文本标记语言推出"Rich HTML article",不断实现科学论文的动态扩充[3]。随后爱思唯尔(Elsevier)公司对海量医学信息资源进行深度标引,方便用户在同一平台获取更全面的资源[4]。Nature 出版集团则通过关联数据提高文献的内容衍生能力,有效帮助用户获得更全面、更有逻辑性的期刊资源[5]。时至今日,国外数字出版内容增强已经获得了一定的成果,各类平台推出的智慧论文[6]、未来论文[7]、泛在论文[8]也成为数字出版内容增强的重要表现形式。国内数字出版内容增强相关研究晚于国外,中华医学会在期刊文章标签集(JATS)的基础上推出中文医学期刊存储交换标准,有效实现该社期刊的资源整合,为中

* 本文系国家社会科学基金重大项目"新时代我国文献信息资源保障体系重构研究"(项目编号:19ZDA345)研究成果之一。
作者简介:翟姗姗,副教授,E-mail:zhais@mail.ccnu.edu.cn;刘星月,硕士研究生;陈欢,硕士研究生。
收稿日期:2021-06-16 修回日期:2021-09-22 本文起止页码:96-107 本文责任编辑:易飞

文全文数字出版迈出了关键的一步[9]；崔玉洁等[10]通过探讨增强出版的传播内容为期刊数字化建设提出新的思路；宋宁远等[11]等对各类增强型出版物的各类模型展开比较分析；朱琳峰等[12]则通过总结国内外数字出版内容增强的相关实践探讨我国学术期刊内容增强的发展策略。

无论是国内还是国外，数字出版内容增强的研究离不开语义问题的探讨。事实上，语义出版是数字出版的高级形态，数字出版内容增强是数字出版向语义出版的重要过渡。因此，在探讨数字出版内容增强相关问题时，引入语义出版是十分有必要的。D. Shotton在2009年首次提出语义出版的概念[13]，王晓光等于2011年将语义出版概念引入我国[14]。随后，苏静等[15]将数字出版与语义出版进行对比，发现数字出版强调科学数据的开放共享、重复利用，而语义出版则更加侧重利用关联数据等技术促进文献内容的富语态化；李楠等[16]通过分析国内外语义出版相关实践，总结出语义出版的技术框架；乐小虬等[17]则在2016年推出了面向语义出版的结构化论文写作工具Dpaper，一定程度上实现了论文在写作阶段的结构化，成为中文语义出版的重要实践。

数字出版经过学者们长期的研究，已经获得了较为成熟的研究结果，但数字出版仍无法解决学术资源呈现单一、缺乏内容深度加工等问题，而语义出版注重深入内容，挖掘语义信息，为数字出版开拓出了新的思路，但值得一提的是，受中文语义、技术难度等限制，直接将语义出版运用于中文文献信息资源建设存在一定困难，因而，面向语义出版的数字出版内容增强成为文献信息资源服务的更好选择。

2.2 医学文献信息资源建设与服务相关研究

文献信息资源是一个集合概念，其范围相当广泛，其中包括以学科特征为划分标准的专业文献信息资源，医学文献信息资源就是其中一个重要的门类。国外对医学文献信息资源的建设实践相对成熟。早在2004年，欧盟药品监督管理局建设了临床试验数据平台，为研究人员进行临床试验提供信息集中共享服务[18]。此外美国生物与生命医学文献资源整合中心对100多个期刊上的医学文献进行了聚合，从而搭建了医学检索的重要平台[19]。而相较国外，国内对医学信息资源建设与服务还处于起步和快速发展阶段。一方面，研究人员认为医学图书馆是医学文献信息资源建设与共享的重要依托，因而以医学图书馆为切口展开探讨，如方芳等[20]人对复旦大学图书馆联盟进行了馆藏资源订购、馆员培训等方面介绍；程鸿等[21]对建设省级医学数字图书馆联盟进行了可行性、必要性分析；另一方面，更多研究人员利用新型技术探讨医学信息资源建设上的可能性，如苏春萍等[22]提出了基于语义网和SOA技术的医学图书馆信息服务模型，为医学专业用户提供医学语义智能化检索服务；张军亮等[23]借助语义关联技术，构建多源信息资源发现系统，为临床医生提供科学研究和临床决策服务；翟姗姗等[24]将知识检索与分面检索结合，构建在线医疗社区分面检索模型。

不难发现，相较其他类别的文献信息资源，医学文献信息资源结构化程度更高，已经取得了一定的研究进展。但仍存在一些明显的问题：①资源分布方面，我国虽然在长期的医疗研究与实践中积累了海量医学文献资源，但这些资源广泛分布在不同医学领域和机构中，跨医院、地区、领域的医学资源系统性建设十分困难，存在较为明显的"信息孤岛"现象；②资源呈现方面，医学文献呈现方式单一，PDF仍是最常见的资源获取格式，受此制约，医学文献之间关联不足，也无法对专业性极强的医学文献展开知识挖掘；③资源描述方面，随着技术的不断进步，不少学者试图借助本体、知识图谱等新型技术优化医学文献服务，但国内仍缺乏统一且专业的医学元数据方案，此外领域本体、公共本体的缺乏也导致医学文献建设缺乏基础条件；④资源服务方面，国内医学资源平台建设不够完善，用户仍需花费大量精力辨析关键信息，这对用户的信息检索、信息识别能力提出了较高的要求。因此本文提出面向内容增强的医学文献信息资源服务框架，希望在一定程度上为中文医学文献信息资源现存问题的解决提供契机。

3 面向内容增强的医学文献信息资源创新服务模式整体思路

内容增强导向下，医学文献信息资源服务产生了一系列新需求，本文在需求导向下设计了医学文献资源出版物内容增强的整体方案。

3.1 面向内容增强的医学文献信息服务新需求

在医学研究与实践的长期探索中，文献信息资源成为凝结医学成果的主要方式。但随着学术研究进程的不断加快，用户更加追求结构鲜明、重点突出、呈现多样、内容丰富的医学文献信息资源，希望在提高资源利用率的基础上尽可能加快医学知识学习进程。而基

于传统数字出版的医学传统出版物多以文字形式呈现,且受学科属性限制,其内容往往生涩难懂,即使是有学科背景的医学工作者,也需要花费大量时间、精力进行医学研究,在漫长的学习过程中,大量有价值的医学文献信息资源被遗漏甚至忽略,医学文献信息资源学习往往事倍功半。相较传统数字出版,经过内容增强的医学增强出版物在内容表达、知识扩展等层面具有显著优势,医学增强出版物主要从以下 4 个方面实现内容增强:

3.2.1 出版物结构化程度增强

长期以来,传统题录项可以满足用户对出版物的基本认知需求,但出版物作为内涵丰富的有机体,用户想要从中获取知识仍需深入复杂的论证过程,这对用户的个人能力提出了较高要求。因此,数字出版内容增强需要从两方面对传统医学出版物进行结构化处理:一方面,医学增强出版物保留传统的元数据出版方式,如关键词、标题、机构等信息;另一方面,在传统元数据出版方式基础上,数字出版内容增强对出版物论证过程展开梳理,将功能结构引入增强出版物,强调问题、假设、方法、数据源、实证、结论等信息,并对其进行标注,帮助用户梳理论证过程,快速定位关键信息。经结构化处理后的医学增强出版物则以最直观的方式为用户呈现有效信息,极大加快了用户医学知识获取的进程。

3.2.2 出版物语义化程度增强

深入内容层面挖掘内容语义是用户实现医学知识获取的关键条件。一方面医学出版物包含诸多内容元素,简单有效地呈现内容要素及其关系是用户快速掌握关键信息的基本前提,因此基于元素的图、表、定义、公式、数据集、算法等信息剖析尤为关键;另一方面,知识获取是用户进行出版物阅读的最终目的,因而医学增强出版物充分挖掘基于知识的医学实体、实体类型、实体属性等信息,并充分利用知识图谱等知识工具辅助用户理解资源,进而促进资源的知识价值最大化。

3.2.3 出版物丰富性程度增强

增强出版从两方面丰富医学出版物,进而为用户提供更加丰富的服务:一方面增强出版物内容更加丰富,不仅包含文字、图表、语音等信息,还会增加用户检索、浏览、下载、借阅、订购、评论等数据信息,有效的日志数据或 UGC 数据使用户个性化服务的获取成为可能;另一方面,医学增强出版物元数据丰富性增强,医学描述元数据在受控词表的基础上,灵活地定义资源结构及其特征,能够有效缓解资源融合过程中的兼容问题,除了描述元数据,医学增强出版物同样蕴含丰富的管理元数据,通过访问与使用、资源反馈等信息更好满足用户的使用需求。

3.2.4 出版物开放关联程度增强

传统出版下,医学出版物所含的数据信息量有限,用户需要通过一定的检索行为来保证知识获取的全面性,因此,将相关资源尽可能多地关联到出版物是十分有必要的。医学增强出版物的开放关联分为显性关联和隐性关联两类:通过显性关联,医学增强出版物关联具有直接链接关系的参考文献、引证文献、引用文献等,以实现延伸内容的转化与应用,满足用户对信息获取和传递的更多需求;通过隐性关联,医学增强出版物基于知识建立实体及其属性间的关联关系,从内部关联和外部关联两个角度实现医学出版物的横向扩充和纵向延伸。

3.2 医学文献资源出版物内容增强的整体方案设计

传统数字出版包含人工干预、结构识别、格式解析等流程,过程较琐碎且最终呈现效果不甚理想[25]。语义出版则主要包含语义元素的识别与描述、语义关系揭示与关联、语义网络的展示与交互 3 个流程。本文对数字出版过程进行简化,同时结合语义出版,提出医学出版物内容增强的整体框架,见图 1。

如图 1 所示,医学文献资源出版物内容增强的整体框架包含实现流程和创新服务两大部分。医学传统出版物经过元素识别、语义映射、语义描述、多维关联 4 步实现数字出版内容增强,其成果被称为医学增强出版物。医学增强出版物则从结构化程度、语义化程度、丰富性程度、开放关联程度 4 个方面实现医学文献信息资源的内容增强。事实上,传统医学出版物缺乏内容的深层处理,用户实质获取到的医学文献服务十分有限,因而用户需要花费大量时间去理解生涩枯燥的专业医学知识,知识转化效率低下,而医学增强出版物从检索、阅读、获取等方面为用户提供更全面、更灵活、更完善的服务,在出版阶段为用户的知识交流打下基础。

4 增强型医学出版物创新服务模式的实现

经过元素识别、语义映射、语义描述、多维关联这一流程,医学传统出版物将转变为增强型医学出版物,从检索、浏览、资源获取 3 方面实现服务机制创新。

4.1 元素识别

语义元素可以反映资源的语义特征,因此需要对资源的语义元素进行挖掘,剖析其内在逻辑关联关系。传统医学出版物往往局限于粗粒度知识单元,为了实

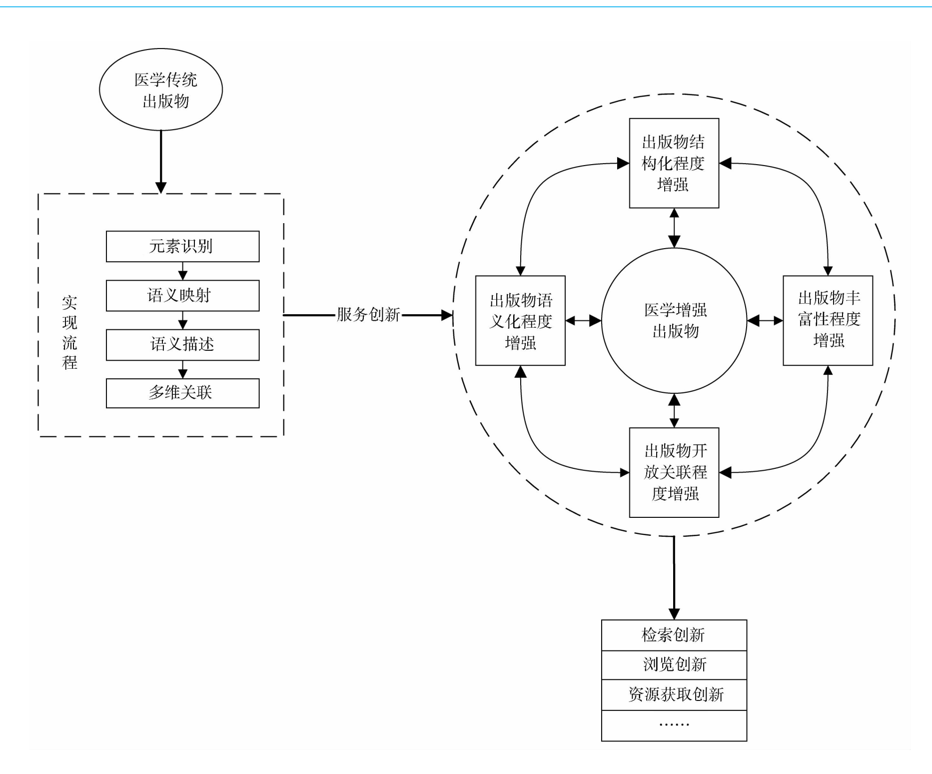


图1　医学文献资源出版物内容增强的整体方案

现医学出版物的内容增强,需要尽可能用最直观的方式呈现出版物蕴含的潜在知识。医学出版物包含丰富且复杂的专业医学内容,因而需要对其进行拆分,从不同碎片中识别出关键元素(见图2),医学内容增强的元素识别从医学出版物的外部特征和语义特征两个方面出发。

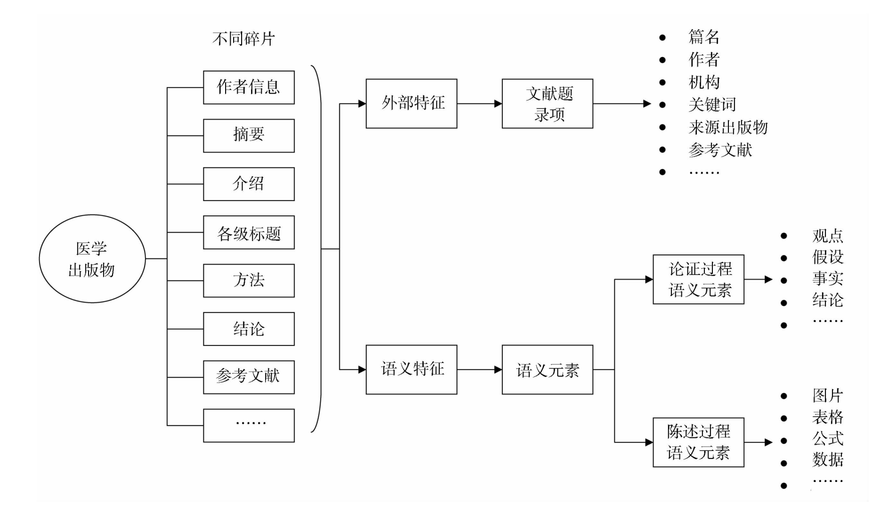


图2　医学增强出版物元素识别

外部特征即传统医学出版物的基本题录项字段,如出版来源、机构、作者等信息。医学出版物外部特征鲜明且容易获得,事实上,在学术期刊长期探索与研究中,题录项外部信息已经成为表达精准的有效字段,因而充分掌握与识别医学出版物的外部特征是十分有必要的,对医学出版物外部特征的识别,也是展开医学文献信息资源内容增强的必要前提;语义特征则需要深入出版物内容层面,从篇章、字词句、语用等多个层面辨析语义元素,通过语义元素来反映医学文献资源的知识内涵,而这些元素组合成了若干知识单元,富含丰富的知识信息,是用户理解与学习医学出版物的要点。医学出版物内容增强通过资源格式转化,统一为XML格式,再通过内容拆分,识别出出版物蕴含的细粒度知识单元,通过标注、突出显示等方式突出关键信息,凸显文献的结构层次。

元素识别是医学出版物内容增强的首要前提。通

过元素识别,直接从源头对复杂的医学出版物进行结构化处理,用户不必逐字逐句进行研读,而可以在医学增强出版物的辅助下直接进行详略得当、主次分明的知识学习,医学增强出版物在一定程度上为用户过滤掉冗余的信息,直接为用户呈现必要的关键信息,让医学研究者展开高效率的深度学习成为可能。

4.2 语义映射

语义映射是医学文献信息资源内容增强的关键步骤,语义映射对拆分出的题录项元素与语义元素展开映射。语义映射需要充分利用CMeSH主题词表,CMeSH即《中文医学主题词表》,是由中国医学科学院医学信息研究所编制的受控医学叙词表,以CMeSH主题词对医学文献信息资源进行描述是生物医学权威文献数据库中组织和表达的重要形式,利用医学规范序词可以有效排除"同词异义""一异多词"等语义含糊现象。值得一提的是,CMeSH通过概念组配来表达文献主题,这些规范序词以概念逻辑为基础,在主题词表的加持下自身便蕴含语义关系,即同义、隶属、相关等语义关系。此外,CMeSH依据学科体系和逻辑关系将字顺表中互不联系的主题词进行逐级排列,按等级隶属关系生成等级结构鲜明的树状结构表,用户可以借助CMeSH主题树进行知识探索。

CMeSH作为规范化的主题词表,其表述与用户经常使用的自然词汇具有一定差距,如自由词"小儿麻痹"对应的主题词为"脊髓灰质炎"。为了实现资源的规范描述,也为了更加充分地利用CMeSH内涵的语义关系,需要实现用户自由词到CMeSH词汇的映射。孙海霞等[26]融合生物医学词汇字长和CMeSH语义关系,提出自由词到CMeSH主题词的语义自动映射方案(见图3)。自由词到CMeSH规范主题词的映射遵循了CMeSH的两个准确性原则:①CMeSH主题词越长,表达的概念越准确;②在CMeSH词表中,距跟节点越远,表达的概念越准确。由此,确定了自由词到主题词映射的基本流程:①利用分词技术,设定好既定阈值,低于阈值的词汇不作为关键词保存,可直接排除;高于阈值的保留,继续进行主题词映射。②当存在唯一的主题词时,直接建立自由词与主题词间的映射关系。③当存在多个主题词,且存在共同上位主题词时,取最近的共同上位主题词(当存在多个最近的共同上位主题词时,则比较路径,选择路径较长者;若路径长相同,选择最长字符串;若两者都相同,选择随机)④当主题词不唯一且不存在共同上位词时,选取相似度最高的最长字符串。本文借助映射方案,将识别出的自由词映射到CMeSH主题词,并将这些规范的词汇作为元数据关键词项存储在数据库中,从而保证医学资源的规范性描述。

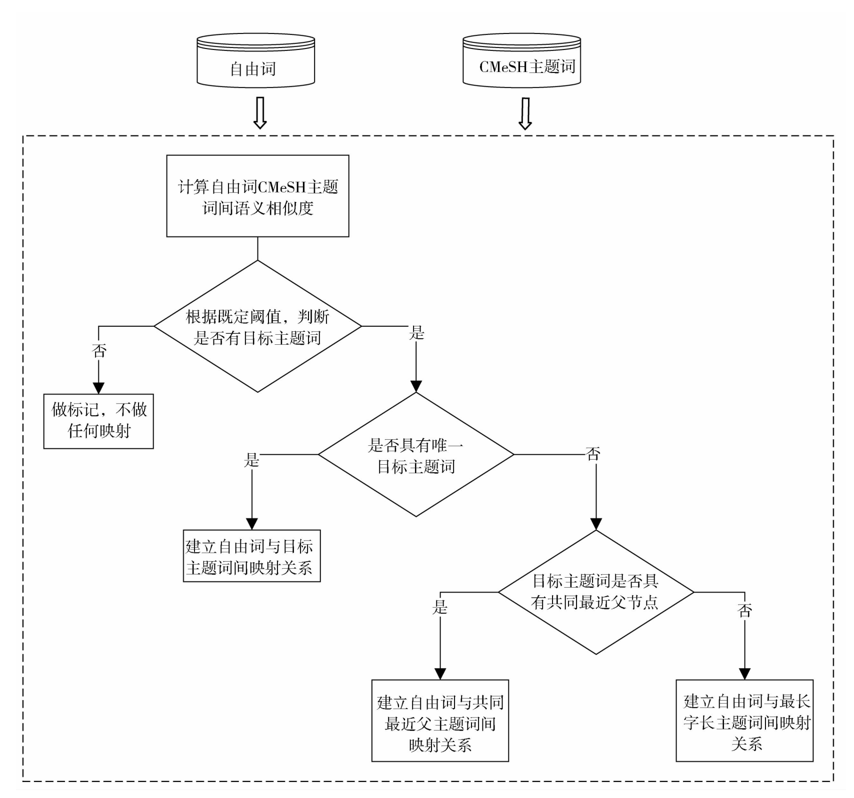


图3 基于CMeSH的自由词-主题词映射方案

4.3 语义描述

语义描述是医学出版物内容增强的中心步骤，元数据相关技术是内容增强语义描述的实现方式。元数据是关于数据的数据，是用来理解并解释资源的数据。通过建立元数据方案，对出版物全文展开语义描述，可以有效辅助医学文献信息资源的开放获取，也可以从微观角度解释语义信息。针对医学文献信息资源，有两类常见的元数据方案，一类是直接使用 DC 元数据，只对限制属性进行扩展；另一类以 DC 元数据为基础，结合 CMeSH 提出新的元数据字段。本文借鉴 MCM (The Medical Core Metadata)元数据方案[27]，该方案直接采用复用 15 个 DC 元数据元素，并结合 CMeSH 词表对部分限制属性进行拓展。此外，本文参考《新闻出版内容资源加工规范》，增设“加工深度标识”这一核心元素，以此标识资源内容的揭示深度，取值为“不加工、粗加工、深加工”，以此对应资源呈现的不同粒度，最终形成完整的医学文献信息资源元数据方案，如表 1 所示：

表 1　MCM 元数据方案

序号	核心元素	扩展	序号	核心元素	扩展
1	题名		9	格式	与 Internet 资源格式一致
2	作者		10	资源标识符	可增加的索取号或 ISSN
3	关键词	Mesh，副标题	11	来源	期刊或项目
4	描述		12	语种	
5	出版者		13	关联(层次)	
6	其他责任者		14	覆盖范围	主要的主题
7	日期		15	权限	版权声明等
8	资源类型	扩展的 Internet 资源类型	16	加工深度标志	

语义描述是多粒度资源展示的基础。传统医学出版物多以全文为单位对资源内容进行展示，粗粒度的资源展示缺乏重点，不利于用户的知识获取，为了满足用户多粒度信息需求，需要借助语义技术改善传统单一粒度的信息组织方式。变粒度医学文献信息资源内容重组基于传统的粗粒度出版方式，增加中粒度、细粒度的知识组织方式。变粒度内容重组将关键词作为最小的知识单元，本文通过 CMeSH 语义映射，将内容蕴含的关键词转换为主题词，进而将所有蕴含相同主题词的资源聚合在一起(见图 4)。通过内容重组可以为用户呈现更加直观的医学信息，变粒度的资源呈现则给用户提供了更多学习方式，用户可以根据自己的需要选择不同呈现方式的资源。

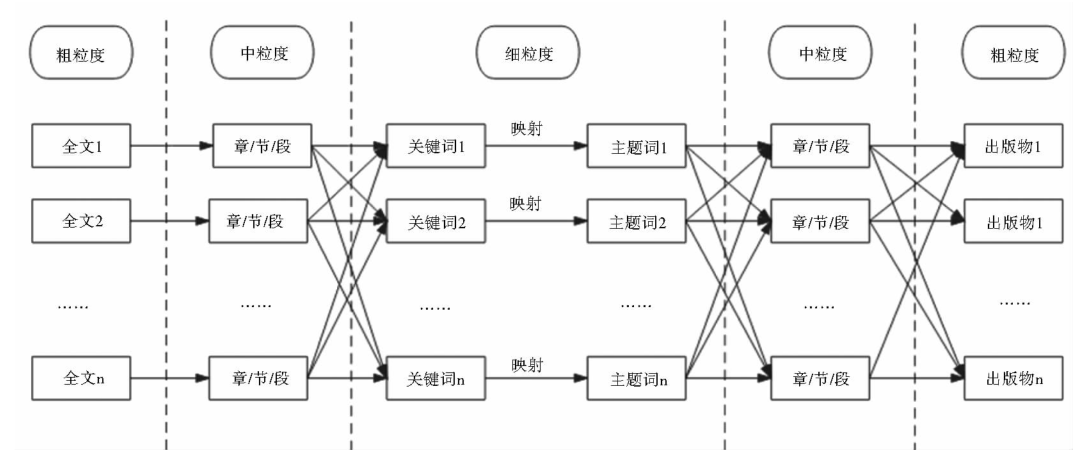


图 4　医学文献信息资源多粒度内容重组

4.4 多维关联

多维关联是医学出版物内容增强的必要流程，其目的在于用有效的外部资源丰富医学出版物，帮助用户在尽可能少的浏览和检索下获取尽可能多的知识。本文利用关联数据技术，从显性关联和隐性关联两个角度实现医学文献信息资源的多维关联。

显性关联指可以直接得到链接关系的关联，如参考文献等，因此无需加工处理，直接利用现有的链接关系展开关联。隐性关联指需要经过加工处理得到的关联关系，分为内部关联和外部关联两种，建立的是实体

及其属性间的关联关系。

其中,内部关联指医学出版物依据基本题录特征实现的资源关联,如作者合作网络、机构共现、关键词共现,目前,传统题录项下的内部关联发展比较成熟,可直接利用知识图谱工具进行构建;外部关联则基于语义或知识元,本文以映射出的 CMeSH 主题词为基点展开关联,其中不同粒度的知识单元可彰显不同层次的语义,因而由语义特征延伸出的语义关联可以是不同粒度资源的交错关联。需要强调的是,外部关联包括医学出版物与网络中其他的关联数据集所建立起的关联。图 5 展示了医学文献与医学相关责任者、医学会议、医学机构、医学项目、医学数据库、医学百科 6 类外部资源的关联,其中实例之间的关联通过类属性来揭示[28]。

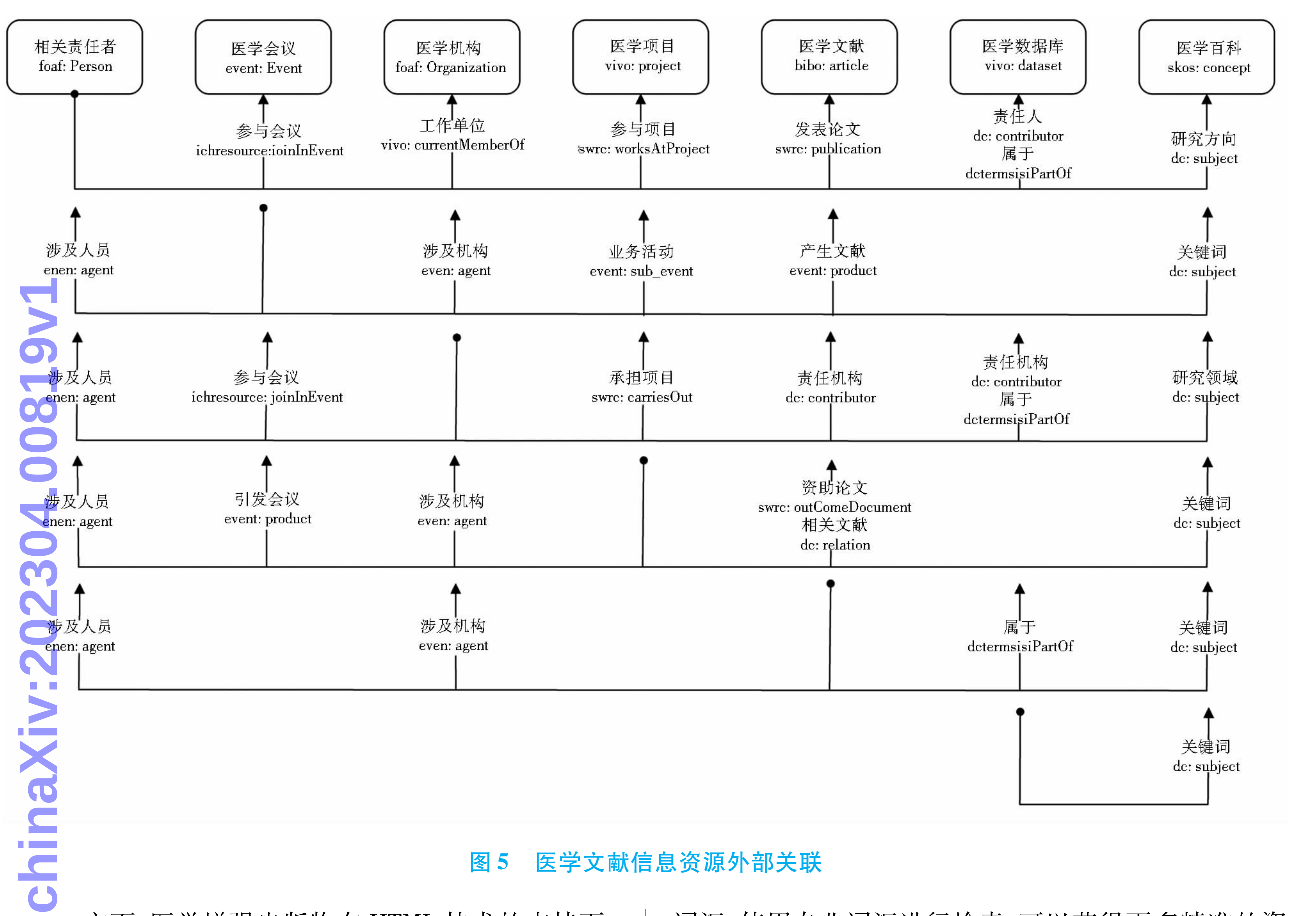


图 5　医学文献信息资源外部关联

一方面,医学增强出版物在 HTML 技术的支持下,具有良好的动态性,因而可以打破传统医学出版物的静态限制,有效实现医学内容的动态扩充。另一方面,增强医学出版物通过多维关联在一定程度上对孤岛壁垒展开冲击,因而一定程度上可以辅助用户展开科学交流。

4.5　医学内容增强平台服务创新机制

知识服务的过程是满足用户需求的过程,满足用户的知识需求也是知识服务的核心。本文针对医学出版物特征,创新性地将 CMeSH 医学主题词表引入出版物内容增强,从检索、浏览、资源获取等方面满足用户多层次的医学知识获取需求。

4.5.1　检索创新

用户可直接在医学内容增强平台进行自由词检索和知识元检索。医学文献信息资源蕴含大量专业医学词汇,使用专业词汇进行检索,可以获得更多精准的资源,但与之相对的是,非专业人士无法掌握有效的专业词汇,这类用户更多使用自由词进行检索。通过本文建立的"自由词(关键词) - CMeSH 主题词"映射,用户直接使用自由词进行检索,便可获得相对精准的医学文献。

4.5.2　浏览创新

医学内容增强平台不仅可以通过突出显示与语义标注突出出版物的关键信息,还可以通过内容重组为用户提供变粒度的资源浏览服务,使用户直接获得变粒度的文献资源、包含检索词的段落章节、与检索词相对应的规范主题词及其诠释,尽可能满足不同用户的不同需求。此外,用户任意输入检索词,平台能够以主题树的方式为用户呈现相关主题词,一定程度上可以启发用户展开医学知识探索,挖掘用户的潜在信息

需求。

4.5.3 资源获取创新

用户通过医学内容增强平台可获取 PDF、HTML 等多格式的资源，也可以基于科学论文获得与之相关的网络健康信息。本文通过关联数据技术展开显性关联和隐形关联，在传统出版物基础上增添多类型的网络健康信息，用户不仅可以获得文本类资源，还可以获得与之相关的图片、数据等资源。值得一提的是，内部关联中的语义关联是基于 CMeSH 主题词展开的关联，不同于传统题录项关联，基于 CMeSH 主题词的语义关联深入内容，从语义层面尽可能地保证用户资源获取的完整性。

5 面向医学增强出版物的儿科文献信息资源服务创新应用

本文在分析儿科医学文献资源特征基础上，将医学增强出版物整体框架的应用于儿科文献信息资源，并对儿科医学资源内容增强平台展开介绍。

5.1 儿科文献信息资源特征分析

儿科是重要的医学领域。从学科研究的角度，婴幼儿童作为生命发展的初期阶段，对于其他医学领域研究具有重要的参考意义，而现有的医学数字平台更多将儿科作为一个普通的学科类目，学界亟需一个更专业便捷的平台去获取专业的儿科知识；从研究对象的角度，儿科的医治对象处于生长发育期，因而涉及诸多基础学科，进而导致相关资源内容繁多、结构复杂，多源异构的儿科文献信息资源迫切需要被重新组织；从用户需求的角度，儿科学的研究和服务对象是不具自主能力且抵抗力相对更弱的婴幼儿童，因而需要监护人对医学知识展开主动学习，然而大部分用户习惯从非结构化信息中寻求所需，大量经过加工处理的结构化的医学文献往往无人问津，因此需要依托更具可读性的儿科医学增强出版物为用户提供服务，让更多用户获得更高质量、更科学的医学信息。综上所述，实现儿科出版物内容增强是十分有必要的。

5.2 儿科医学增强出版物服务创新的实现

本文使用 MCM 元数据实现儿科医学资源的语义描述，而元数据是用以表征知识的一种特殊数据，任何被标志的元数据都可通过 RDF 加以实现，RDF 使用 XML 语法和 RDF Schema 将元数据描述转化为数据模型，为元数据互操作提供支持。RDF 作为较成熟的数据技术，在生物医学领域也得到了较广泛的应用，网络上也存在很多大规模的医学 RDF 数据集。因此，为保证医学信息描述的专业性、统一性，本文尽可能地复用现有的医学 RDF 三元组数据集 PubMed，但需要说明的是，该数据集蕴含医学和生物科学领域近 5 亿个三元组，本文只截取儿科领域，在此基础上，为更好满足儿科资源的描述需求，本文根据上文阐述的 MCM 医学元数据方案，通过用户自定义方式规范、拓展了一部分词汇。

依据上文的医学文献信息资源多维关联构建思路，儿科文献信息资源关联数据的构建也从显性、隐形两层面入手。本文在充分建立资源显性关联的基础上，更多深入到资源的内容层面，围绕映射出的 CMeSH 主题词挖掘文本资源间语义关联。本文提取 CMeSH 词表中儿科相关的部分，包括儿科疾病、诊断治疗、学科组织 3 类主题词，图 6 展示了部分的儿科疾病类主题词关系图，平台借助前文构建的自由词－主题词映射关系，将文本以主题词的形式表示，随即依据主题词内涵的语义关系实现文本资源之间的语义关联。

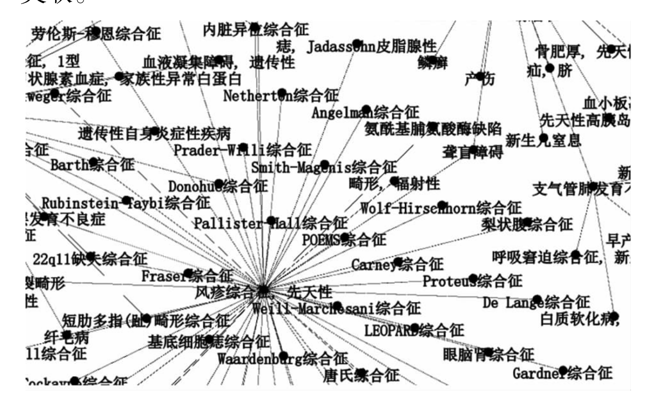


图 6 儿科疾病关系图（部分）

5.3 儿科内容增强平台的创新服务功能展示

为了更形象地展示医学内容增强出版物的各项功能，本文具体到儿科医学文献信息资源领域，搭建儿科内容增强平台，帮助用户从不同维度对医学语义进行查询、浏览、学习。受知识产权等因素的影响，本文选取了《中国实用儿科杂志》《中国小儿急救医学》等 6 个儿科类医学期刊，暂不考虑非儿科期刊中的医学文献，具体见图 7。

首先，儿科内容增强平台可为用户直接提供自由词检索服务。医学词汇专业化程度高，普通用户常使用自由词进行检索，但通常情况下，使用自由词检索往往结果不够精准。平台通过自由词－主题词映射，将用户使用的自由词自动转化为主题词，直接为用户呈



图 7　儿科内容增强平台数字资源展示

现相应的主题词检索结果。如图 8 所示，输入“小儿麻痹”，平台依据映射得到相对应的“脊髓灰质炎”。事实上，CMeSH 主题表包含的不仅是规范序词，还有规范序词之间蕴含的语义关系，如图 9 所示，平台对主题词蕴含的语义关系通过主题树的方式进行呈现，帮助用户跳出思维局限，主动为用户提供更多相关主题词，因而用户不仅可以了解相关知识体系，还可以使用相关的主题词获得更多所需的信息。



图 8　平台自由词 – 主题词对应情况

其次，儿科内容增强平台通过内容重组可以为用户呈现多粒度的文献信息资源。粗粒度的文献展示是平台的基础功能，中粒度、细粒度的儿科文献信息资源呈现则为用户的知识获得提供了更多可能性。如图 10 所示，平台以片段为单位，对检索结果进行中粒度呈现，这些片段全部来自平台蕴含的文献数据库，将所有蕴含相应主题词的片段呈现给用户，片段中的相应主题词也会被标红显示，相较粗粒度的文献展示，用户可以更加直接地获得有效信息。此外，平台通过内容重组深入到文献语义层面，对资源进行细粒度的知识呈现，如图 11 所示，平台从定义、特征、诊断等 10 个方面对儿科信息资源进行重新组织，直接从平台包含的数据库中挑选出关键信息呈现给用户，大大减轻了用户从海量资源获取所需知识的压力。

最后，儿科内容增强平台将传统出版物与网络健康信息相融合，实现传统出版物的内容增强。如图 8 所示，检索“小儿麻痹”，不仅可以获得关于“脊髓灰质炎”的百科解释，还可以获得与之相关的网络图片。再如图 11 所示，平台对“脊髓灰质炎”进行了知识重组，将网络上与“脊髓灰质炎”相关的治疗医院、医生、相

翟姗姗，刘星月，陈欢. 面向出版物内容增强的医学文献信息资源服务模式创新[J]. 图书情报工作,2022,66(6):96－107.



图 9 “小儿麻痹”主题树

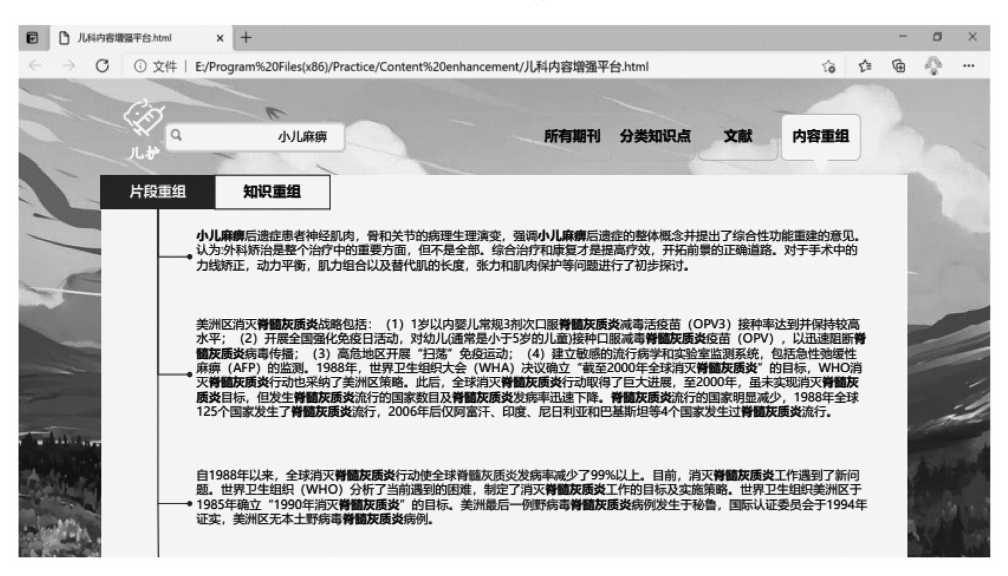


图 10 “小儿麻痹”检索结果中粒度呈现(部分)



图 11 “小儿麻痹”检索结果细粒度呈现(部分)

关视频一起呈现给用户,在文字基础上实现多类型资源的内容聚合,从而为用户提供更加全面的知识信息。

6　结语

本文为了解决医学文献信息资源现存的实际问题,也为了给用户提供更好的医学文献信息资源服务,提出了基于数字出版、面向语义出版的医学出版物内容增强方案,并以儿科文献信息资源为实例,搭建儿科出版物内容增强平台,展示了语义环境下医学出版物内容增强的可行性。

当然,本文的研究只是一个应用方案,搭建的儿科内容增强平台不够成熟,仍缺乏实证效果的反馈测量;另外,语义环境下,各类技术不断进步,知识图谱等技术可深入到知识单元并直接挖掘资源的知识关联,AR等可视化技术则可以更加生动形象地呈现医学知识,因而医学出版物内容增强仍需不断利用新型技术实现医学知识挖掘与利用;最后,开放网络环境下,外部医学资源的获得并非难事,但受版权、地域等原因的制约,外部资源的全面获取仍是一个难题,仍需要各医学出版商给予支持。本文为医学增强出版物的构建与实现提供了一个可供探索的实现方案,希望在医学增强出版物辅助下,用户可以获得更好的医学信息服务。

参考文献:

[1] 刘兹恒,涂志芳. 学术图书馆参与数字出版的动因与条件分析[J]. 图书情报工作,2016,60(3):32-37,113.
[2] 许鑫,江燕青,翟姗姗. 面向语义出版的学术期刊数字资源聚合研究[J]. 图书情报工作,2016,60(17):122-129.
[3] RSC[EB/OL]. [2021-09-18]. https://pubs. rsc. org/.
[4] Elsevier[EB/OL]. [2021-09-18]. https://www. elsevier. com/.
[5] Nature[EB/OL]. [2021-09-18]. https://www. nature. com/.
[6] Smart article [EB/OL]. [2021-09-18]. http://as. wiley. com/WileyCDA/Section/id-817760. html.
[7] ALBERSBERG I J, HEEMAN F, KOERS H, et al. Elsevier's aticle of the future enhancing the user experience and integrating data through applications[J]. Insights: the UKSG journal, 2012, 25(1):33-43.
[8] Anywhere article[EB/OL]. [2021-09-18]. http://exchanges. wiley. com/blog/tag/anywhere-article.
[9] 沈锡宾,李鹏,王红剑,等. 中华医学会系列期刊全文电子文档交换和存储标准初探[J]. 中国科技期刊研究,2015,26(5):475-479.
[10] 崔玉洁,包颖,廖坤. 全媒体出版中增强出版的模式研究[J]. 编辑学报,2018,30(1):70-73.
[11] 宋宁远,王晓光. 增强型出版物模型比较分析[J]. 中国科技期刊研究,2017,28(7):587-592.
[12] 朱琳峰,李楠. 学术期刊数字出版内容增强模式探索[J]. 编辑学报,2019,31(4):421-423,427.
[13] SHOTTON D. Semantic publishing: the coming revolution in scientific journal publishing[J]. Learned publishing, 2009, 22(2):85-94.
[14] 王晓光,陈孝禹. 语义出版的概念与形式[J]. 出版发行研究,2011(11):54-58.
[15] 苏静,曾建勋. 国内外语义出版理论研究述评[J]. 中国科技期刊研究,2017,28(1):33-38.
[16] 李楠,孙济庆,马卓. 面向学术文献的语义出版技术研究[J]. 出版科学,2015,23(6):85-92.
[17] 乐小虬,王子璇,张晓林,等. DPaper:一种面向语义出版的结构化论文写作工具设计与实现[J]. 现代图书情报技术,2016(11):76-81.
[18] 掌握北美临床试验数据中心的检索方法是向 SCI 期刊投稿的第一步[J]. 中国组织工程研究,2012,16(7):1306.
[19] 李娇,寇远涛,黄永文,等. 国内外语义出版实践研究[J]. 数字图书馆论坛,2017(12):25-31.
[20] 方芳,应峻,陈怡,等. 复旦大学医学图书馆资源共建共享联盟的实践探索[J]. 中华医学图书情报杂志 2017,26(6):30-32.
[21] 程鸿,李红军. 省级医学数字图书馆联盟建设研究[J]. 图书馆工作与研究,2012(10):38-40.
[22] 苏春萍,郭喜娟. 基于语义网和 SOA 的医学图书馆信息服务模型设计[J]. 中华医学图书情报杂志,2015,24(9):46-49.
[23] 张军亮. 基于语义关联的多源医学信息资源发现服务系统研究[J]. 图书情报知识,2019(3):113-122.
[24] 翟姗姗,潘英增,胡畔,等. 基于医学知识图谱的慢性病在线医疗社区分面检索研究[J]. 情报理论与实践,2021,44(1):195-203.
[25] 陈敬宇,张阿源. 基于资源重构的数字出版流程设计与评价[J]. 中国出版,2017(19):51-54.
[26] 孙海霞,李军莲,李丹亚,等. 基于 CMeSH 语义系统的领域自由词-主题词语义映射研究[J]. 现代图书情报技术,2013(11):46-51.
[27] 韩夏,张晓林. 描述医学资源的元数据方案[J]. 图书情报工作,2003(12):56-59,23.
[28] 翟姗姗,许鑫,夏立新,等. 语义出版技术在非遗数字资源共享中的应用研究[J]. 图书情报工作,2017,61(2):23-31.

作者贡献说明:

翟姗姗:论文指导与框架修改;
刘星月:论文撰写与修改;
陈欢:实证数据采集与管理。

Innovation of Medical Literature Information Resource Service Modes Oriented to the Enhancement of Publication Content

Zhai Shanshan　Liu Xingyue　Chen Huan

School of Information Management, Central China Normal University, Wuhan 430079

Abstract: [Purpose/significance] Medical literature information resources are the necessary prerequisite for users to learn medical knowledge and carry out medical research. However, the existing medical literature information resources have the problems of unstructured and single presentation form. Therefore, the content enhancement scheme of medical publications is proposed in order to provide users with better services of organization and utilization of medical resources. [Method/process] Oriented by semantic publishing, CMeSH medical subject words were integrated into element recognition, semantic mapping, semantic description and multi-dimensional association processes to realize the transformation from traditional medical publications to medical enhanced publications. At the same time, this paper constructed content enhancement examples of medical publications with pediatrics as the application background and verified the feasibility of digital publishing content enhancement in medical literature resource services. [Result/conclusion] It is found that through a series of processes of content enhancement, traditional medical publications can be transformed into medical-enhanced publications, thus optimizing the user learning process of medical knowledge and making it possible for users to learn efficiently.

Keywords: content enhancement　digital publishing　medical literature information resources　CMeSH　semantic publication